

13. The Influence of the Sorites Paradox in Practical Philosophy

[Forthcoming in S. Oms and E. Zardini (eds.), *The Sorites Paradox* (Cambridge University Press); please cite that version]

Hrafn Asgeirsson

In this chapter, I discuss some of the main ways in which the Sorites Paradox is relevant to practical philosophy.¹ I begin by distinguishing between two types of roles that sorites arguments play in the recent literature – one indirect and one direct. The most prominent indirect role that such arguments play is a demonstrative one; sorites arguments are used to show that a particular predicate has borderline cases, from which something of philosophical interest is supposed to follow. It has been argued, for example, that if moral predicates have borderline cases, then most – if not all – forms of moral realism must be false. It has also been argued that the fact that legislation is riddled with vague predicates shows that we need to revise the standard notion of the Rule of Law.

When sorites arguments are employed in a more direct way, they are typically used to show either that the first item in the sorites series is practically problematic or that tolerance is – i.e. the paradox is seen as a motivation to deny one or more premises in the reasoning, rather than, say, deny the validity of the argument or bite the bullet vis-à-vis its conclusion.² The paradox has been used, for example, as the basis for a slippery slope argument against abortion at any time, arguing soritically from its permissibility at some time t_0 after conception to a seemingly absurd conclusion of its permissibility at some time t_n – say, just prior to birth. Here, the ‘solution’ is to reject the first step in the reasoning. The paradox has also been used as a basis for criticism of several important arguments that in one way or another involve the transitivity of value – including, for example, an argument against the adequacy of orthodox decision theory. In these cases, the ‘solution’ is to reject the inductive step, i.e. to reject tolerance.

¹ Note that I will assume a (very) basic familiarity with the paradox in the following discussion. For an overview of the basic notions involved, see Oms and Zardini (this volume), esp. sec. 1.3.

² For an overview of the main solutions to the paradox, see *ibid.*, sec. 1.4.

The first part of the chapter surveys some of the ways in which the Sorites Paradox has figured in arguments in practical philosophy in recent decades, with special attention to arguments in which the paradox plays a more direct role. Given the limitations of space, some significant work – much of it recent – cannot be covered here, but hopefully the illustrative discussion will still be informative, if not complete. To highlight what is at stake, I will outline some of the most prominent arguments in the literature, all of which involve the transitivity of value in some way: Ruth Chang’s ‘chaining’ argument for the novel value-relation *parity*, Warren Quinn’s ‘self-torturer’ argument against the adequacy of orthodox decision theory, and Larry Temkin’s ‘continuum’ argument (later ‘spectrum’ argument) against the transitivity of *better-than*.³

The second part is slightly more probative, focusing on two main themes. First, I further address the relationship between the Sorites Paradox and the three arguments mentioned above, by elucidating in what sense they rely on (something like) tolerance principles. Second, I briefly discuss the prospect of rejecting the respective principles, aiming to show that we cannot do so for Quinn’s and Temkin’s arguments, since – unlike in genuine sorites scenarios – the principles do not function as independent premises in the reasoning but, rather, follow from certain fundamental features of the relevant scenarios. I also try to further distinguish between these two arguments and Chang’s chaining argument by showing that not even adopting what is arguably the most radical way to block the Sorites Paradox – that of weakening the consequence relation – suffices to invalidate the former two arguments. They may of course be problematic in other respects, but if such a radical solution to the paradox does not block them, then at least we have very strong reason to believe that they are not genuinely soritical.

Weakening the consequence relation invalidates Chang’s argument, however. It doesn’t immediately follow that the argument is soritical, but it does mean that – on the account under consideration – it fails for the same reasons as the Sorites Paradox. In both cases, logic has been ‘pushed too far,’ so to speak; chaining together simple, ‘locally’ valid, argument forms, ends up producing an invalid, complex argument form.

³ See Chang (2002), Quinn (1990), and Temkin (1996), respectively.

1. The Sorites Paradox and practical philosophy: Some examples

The following overview of the influence of the Sorites Paradox in practical philosophy is by necessity both selective and fairly brief. As I mentioned, I want to try to impose a bit of structure on the discussion by distinguishing between two types of roles that sorites reasoning plays in the literature. One is indirect, in the sense that Sorites reasoning plays a part in establishing that a particular predicate has borderline cases, which in turn has certain purportedly important implications for the domain in question. The other role of the paradox is more direct, in that certain important arguments in practical philosophy are charged with being soritical, and fault is found with the respective tolerance principles.

Although this twofold categorization is a useful starting point, we should not attach too much importance to it. As we will see even from this selective overview, the variation within each category is so great that the distinction between indirect and more direct applications of sorites reasoning is ultimately not all that informative. And, as is to be expected, many of the interesting bits tend to be related to the particular domains of application. Nevertheless, the discussion here should still provide clear indication of how developments in philosophy of language and philosophical logic, generated specifically in response to the Sorites Paradox, may have significant implications for several important arguments in practical philosophy.

1.1.1 Indirect uses of sorites reasoning: The debate about moral realism

One of the areas of practical philosophy in which the Sorites Paradox has been most influential is the recent debate about moral realism.⁴ Here, the paradox is typically used demonstratively to show that moral predicates have borderline cases, the existence of which present a large *prima facie* problem for the idea that the truth of moral judgments is independent of the attitudes associated with them – i.e. that such judgments report facts.

As Shafer-Landau (1995) points out, it is easy enough to see why moral realism might be thought to be at odds with vagueness. Moral realism may seem to naturally favor *epistemicism* – i.e. the view that in each borderline case, there is a fact of the matter whether the relevant predicate applies, albeit an unknowable one.⁵ Such irremediable ignorance can

⁴ For a general overview of moral realism, see e.g. Sayre-McCord (2017).

⁵ For what has become the classic exposition and defense of epistemicism, see Williamson (1994). See also Magidor (this volume).

also – on this view – account for our lingering disagreements about moral matters and our doubts about what to do in moral dilemmas: These phenomena are persistent because the ignorance in question cannot even in principle be resolved.

However, Shafer-Landau says, if epistemicism about moral vagueness is true, we end up with implausibly many unknowable moral truths, given the prevalence both of seemingly irresolvable disagreement about moral matters and of cases in which competing values appear to fail to resolve the issue at hand. Irrespective of the plausibility of epistemicism in other domains, he thinks, moral realism carries with it some commitment to significant epistemic access to the relevant facts (if they exist).⁶

Now, on a simplistic view of the available options, this could easily seem to motivate a rejection of moral realism. If it seems implausible that morality is fully determinate yet ubiquitously unknowable (as epistemicism about moral vagueness may seem to entail), then one may wish to try to explain moral disagreement *without* reference to moral properties. Just as some have argued that the lesson to be learned from the Sorites Paradox is that there are no heaps, bald men, or clouds, so too a non-cognitivist can claim that the paradox shows that there is no generosity, bravery, or goodness. In short, the response here to the implausibility of epistemicism – either wholesale or specifically in the domain of morality – is to embrace *nihilism* (again, either wholesale or specifically with respect to moral properties): If there are no precise moral properties, then there are no moral properties at all.⁷ Or so this line of reasoning goes.

As Shafer-Landau correctly points out, however, such a move is unnecessary. We can reject both epistemicism and nihilism and retain moral realism even in the face of widespread irresolvable disagreement and conflict of values. Shafer-Landau's preferred view is one on which we can make sense of *vague properties* – i.e. real properties that have 'indeterminate extensions,' as he calls it.⁸ A property P has an indeterminate extension, on this view, if, for some object *o* (or set thereof), there is no fact of the matter (and so no truth

⁶ See also Schoenfield (2015) for further arguments against epistemicist explanations of the vagueness of moral predicates.

⁷ Or such properties are necessarily uninstantiated. Thanks to Elia Zardini for helpful comments on this point. For a wholesale approach to nihilism, see Unger (1979); see also Eklund (this volume), sec. 6, and Kurtzal (this volume), sec. 3.

⁸ Shafer-Landau (1995), 84 and 93.

about) whether *o* instantiates P. In the moral domain, for example, there may be no fact of the matter whether a person, or an action, instantiates generosity.

At least *prima facie*, there are certainly other accounts of vagueness available to the moral realist – in particular, so-called *semantic* accounts, like *contextualism* and *supervaluationism*. Many authors argue, however, that purely semantic explanations of moral vagueness are problematic.⁹ The reason is that such accounts generally posit what has come to be known as a *shifty* semantics for vague predicates, which doesn't sit well with what we tend to consider as acceptable ways to resolve moral borderline cases. Put very briefly, the worry is that – in some significant sense – purely semantic views allow for borderline cases to be resolved by linguistic considerations alone. If that is correct, then such accounts have – at least from a moral realist perspective – counterintuitive implications for how we can resolve vagueness-related moral quandaries. Intuitively, a *real* solution to such quandaries requires some reference to reasons, rather than merely choosing to use our words differently.¹⁰ Consequently, these authors think that moral predicates require a *rigid* semantics, on which semantic vagueness is grounded in *ontic* vagueness. Taking moral vagueness seriously, then, arguably requires locating the vagueness 'in the world.' Or so this line of reasoning goes. The properties 'in the world' that Shafer-Landau – specifically – has in mind are non-natural, non-reducible moral properties, constituted by natural properties.¹¹

Constantinescu (2013), however, argues that the moral realist has fewer options than one might think, insofar as we want to be able to take moral vagueness seriously. In particular, he argues that ontic moral vagueness is not compatible with non-naturalist moral realism like Shafer-Landau's. The reason, he says, is that it is plausible to suppose that vague moral properties are generally multiply realizable by a very wide range of underlying properties, which entails – on his view – that such properties are essentially disjunctive. Insofar as we are supposed to take non-natural moral properties to be constituted by natural properties, as Shafer-Landau suggests, this gets non-naturalism into trouble, Constantinescu

⁹ See e.g. Shafer-Landau (1994), Constantinescu (2013), and Schoenfield (2015).

¹⁰ Arguably, moral realists are also committed to a view on which moral propositions are – if true – true independently of our attitudes towards them (or towards the associated facts), which may seem to conflict with the conversational discretion afforded by shifty semantics. I think there are ways around this worry, but it is easy enough to see why this is likely to make moral realists uneasy with explanations that rely on such semantics.

¹¹ See e.g. Shafer-Landau (2003). For a more general overview of moral non-naturalism, see e.g. Ridge (2014).

argues. The root of the problem, he thinks, is that there just doesn't seem to be anything that sufficiently unites the (open-ended) set of natural disjuncts in such a way to justify the claim that moral properties are real (insofar as the existence of disjunctive properties turns on similarities between disjuncts). Constantinescu's concern, then, is that due to the vagueness of moral properties there is – on the non-reductive view – no underlying natural property, or set thereof, which is capable of constituting any moral property.¹²

This is a general problem for moral realism, of course, but one that the *naturalist* is able to get around by appealing to other considerations to justify the claim that moral properties exist, such as causal efficacy. Since non-natural properties are generally taken to be causally inert, such considerations are not available to the moral non-naturalist, however.¹³ Or so Constantinescu argues. If that is correct, then it seems that taking moral vagueness seriously within a realist framework favors moral naturalism.

1.1.2 Indirect uses of sorites reasoning: Arbitrariness and the Rule of Law

In philosophy of law, sorites-susceptibility has also been used indirectly to argue that we need to revise the standard notion of the Rule of Law. Unlike in the debate about moral realism, which centers on the vagueness of *moral* predicates, nothing here hangs on the vagueness of *legal* predicates, as such. Rather, the mere fact that legislation is riddled with predicates that admit of borderline cases shows that some legal cases cannot be resolved in a principled way, which introduces arbitrariness into the law – a known adversary of the Rule of Law.

Endicott (2000) has us imagine a set of identical defendants, for example, all of whom must – by law – be prosecuted 'within a reasonable time'.¹⁴ Now, a single day is not going to make a difference to whether or not someone is or isn't prosecuted within a reasonable time, but for some pair of identical defendants within the borderline region the law cannot but prosecute one yet dismiss the charges against the other (insofar as the law

¹² My own response to Constantinescu's concern is that vagueness is not really doing any specific work here. If moral properties are multiply realizable in the way he suggests, then they are so not in virtue of being vague but – rather – in virtue of being (incommensurately) multidimensional. For a discussion of the relationship between these two semantic phenomena, see e.g. Asgeirsson (2015). Broad multiple realizability remains a challenge for non-reductive non-naturalist views, but this is a general metaphysical problem, not specifically related to moral vagueness.

¹³ For a discussion, see e.g. Enoch (2017).

¹⁴ Endicott (2000), 188.

wishes to avoid prosecutions that clearly violate procedural rules).¹⁵ As a result, Endicott says, the law (i) fails to treat like cases alike, (ii) promotes discretion on behalf of legal officials, and (iii) lacks predictability, all of which Endicott takes to be inconsistent with what is traditionally considered fundamental Rule of Law qualities.¹⁶

Since vagueness is – on Endicott’s account – not only in fact pervasive in law, but also a *necessary* feature of it, he takes this to present a very significant challenge to the standard notion of the Rule of Law. The ideal, therefore, must be reconstructed.¹⁷ It is of course possible to regard vagueness in the law as a necessary evil, and to hold that the ideal state of affairs is one in which vagueness in the law is limited as much as possible. However, Endicott points out that minimizing vague law may very well take us *further* away from the Rule of Law ideal, as precision may also introduce arbitrariness by failing to reflect ‘the reasons on which a law ought to be based.’¹⁸ Abandoning the ‘reason of the law,’ Endicott thinks, is certainly inconsistent with the Rule of Law, properly understood, but vagueness does not necessarily entail this; it *may* involve doing so, of course, but so may precise regulation. And sometimes, he concludes, the ability of legal officials to treat like cases in different ways, based on the underlying justification of the relevant law(s), significantly outweighs the costs associated with (i)-(iii) above.

Having given these two examples of indirect uses of sorites reasoning, I want to now move on to more direct applications. Slippery slope arguments sometimes make positive use of sorites reasoning, but the paradox is mainly used critically, to undermine potentially significant arguments of various sorts, each of which involves the transitivity of value in some way. These critical applications will be the focus here. First, however, the sorites series as a slippery slope.

¹⁵ Note that Endicott’s argument relies on a tolerance principle for the term ‘within a reasonable time’. Since most theorists of vagueness reject tolerance principles, this is problematic. However, Endicott’s main point could (more or less) be made by appeal to borderline cases instead of tolerance. Thanks to Elia Zardini for helpful comments here.

¹⁶ See Endicott (2000), 188–9.

¹⁷ See *ibid.*, 189ff.

¹⁸ *Ibid.*, 192.

1.2.1 Direct uses of sorites reasoning: Slippery slopes

The classical example of sorites reasoning used in slippery slope arguments involves arguing from the permissibility of abortion at some time t_0 after conception to a seemingly absurd conclusion of its permissibility at some time t_n – say, just prior to birth. Here, the ‘solution’ is to reject the first step in the reasoning: Abortion at time t_0 after conception is impermissible.

The ‘solution’ crucially depends on treating the argument as valid but unsound – hence the strategy to reject the initial premise (and, importantly, to maintain tolerance). This strategy, however, tends not to be considered a viable general strategy for dealing with the paradox. And even if it were, it doesn’t work for individual claims alone – rather, rejecting for example the initial premise in the argument above would really amount to embracing the nihilist conclusion that *nothing* is permissible.¹⁹ That’s definitely not what proponents of slippery slope arguments are after.

The main problem for slippery slope applications of sorites reasoning, then, is that any feasible response to the paradox will either (i) reject tolerance (i.e. some step in the reasoning other than the initial premise), (ii) reject the validity of the argument, or (iii) bite the bullet and accept the paradoxical conclusion(s). So whichever way we go, any slippery slope argument relying on a sorites series in any robust sense will end up being problematic in some fundamental way. It doesn’t follow, however, that there are no respectable versions of slippery slope arguments involving something like a sorites series, but what will be doing most of the work in such arguments are empirical considerations about the likely consequences of accepting certain morally relevant judgments or rules.²⁰ As a result, these arguments are not genuinely soritical.

¹⁹ That is, this ‘solution’ really amounts to holding that – like other vague predicates – the predicate ‘is permissible’ is *empty*; see e.g. Keefe and Smith (1996), 12–13. Note that the nihilist strategy applies to all moral predicates, including ‘is obligatory’. This may seem – by way of standard duality – to entail that *everything* is permissible. Space is too limited for me to go into this matter more substantially, but the best way to explain these seemingly contrary commitments of the nihilist strategy – I believe – lies in the distinction between *strong* and *weak* permission, where the latter is defined simply as the absence of obligation and the former as something above and beyond that. What the moral nihilist is committed to, then, is, on the one hand, that nothing is strongly permissible and, on the other, that everything is weakly permissible, which is entirely consistent. For a discussion of strong versus weak permission, see e.g. von Wright (1963), 86–87, Alchourrón & Bulygin (1971), Chapter 7, and Navarro & Rodríguez (2014), 78–80.

²⁰ See e.g. Williams (1995).

1.2.2 Direct uses of sorites reasoning: Critical arguments

As I mentioned, the main direct application of the Sorites Paradox is – perhaps unsurprisingly – critical. And in fact, several important arguments in the recent literature in practical philosophy depend on being able to resist the charge that they ultimately depend on soritical reasoning. I cannot discuss all of them here, but will outline some of the most prominent ones, to highlight what is at stake.

Chang's chaining argument for parity

Chang (2002) argues that in addition to the relations of *better-than*, *worse-than*, and *equally-good-as*, there is a fourth value-relation: *is-on-a-par-with*, or *parity*. Consider Mozart and Michelangelo.²¹ We start by assuming that we cannot find a common scale on which we can measure the degrees to which the two artists are creative, although we can count both as very creative. We therefore think that it is false that ‘either one is better than the other or they are of equal value,’ with respect to creativity.²² Chang’s motivating idea for the parity relation is that we would presumably still accept the claim that the imaginary painter Talentlessi – who happens to be *very* bad, and so is at the opposite end of the spectrum relative to Michelangelo – is less creative than Mozart. This generalizes, in the sense that, for any relata r and r' , r and r' can be compared with respect to a feature F if r is a ‘nominal’ instance of F and r' is a ‘notable’ instance of F , or vice versa, *even if* there exists no natural common metric with which we can measure the F ness of r and r' . If this is right, absence of a natural common metric does not entail incomparability. And since we can compare Mozart and Talentlessi with respect to creativity, we can do the same for Mozart and Michelangelo, on the assumption that comparability is preserved through successive, small enough differences in the relevant quality underlying our judgments of creativity.²³ But since we have already accepted that none of the three standard comparative relations applies to the pair, we need the relation of parity to account for such cases.²⁴ Or so the argument goes.

²¹ See Chang (2002), 673–5.

²² See Raz (1986), 324.

²³ Chang (2002), 674; Chang calls this the *Small Unidimensional Difference Principle*. The justification for the principle, she says, is its ‘deep intuitive appeal;’ see *ibid.*, 675. I discuss this principle below.

²⁴ In Chang’s view, parity may explain a wide range of cases – in addition to Mozart/Michelangelo, she mentions a number of other relevant pairs: a career in accounting vs. a career in skydiving, an afternoon at the museum vs. hiking in the woods, and a duty to keep promises vs. a requirement to avoid unnecessary pain; see Chang (2002), 659. Note that we might of course – contra Chang – take

Quinn, and Tenenbaum and Raffman, on the Puzzle of the Self-Torturer

Quinn (1990) presents us with the *puzzle of the self-torturer*, in order to put pressure on one of the fundamental ideas of orthodox decision theory – that the perfectly rational agent should ‘see every moment as a possible new beginning in their practical lives,’ as he puts it.²⁵

Consider a person who is fitted with a device that can deliver a continuous range of electric shock. The device has 1001 settings, from 0 to 1000, and the difference between each pair of adjacent settings is so small as to be imperceptible. It is first set to 0, accompanied with the offer to increase the setting by 1 and receive \$10,000. If the person chooses to increase the setting, she is then presented with the same offer. The puzzle is that for each offer, accepting it maximizes expected utility, yet at 1000 the person would gladly pay the money she has gained for the setting to be returned to 0.²⁶

Quinn identifies the above-mentioned prescription to maximize expected utility at each choice point as the culprit; we should instead pre-select a stopping point and stick to it. As a more general practical matter, he thinks, such a strategy will help us draw the line, for example, between one more bite and too many bites, between puffs of pleasant smoking and lung cancer, or between pleasurable moments of idleness and wasted lives.²⁷ What this means, Quinn thinks, is that we have to supplement the orthodox view of rational decision making with what he calls a ‘quasi-deontological’ element – that of being *bound* in our future choices.²⁸ This will allow agents with non-transitive preferences to act rationally by pre-selecting a stopping point, thereby avoiding the self-torturer’s predicament. But in doing so, we give up what he calls the *Principle of Strategic Readjustment*. Once the agent has picked a stopping point, it is – contra orthodoxy – not rational for her to change her mind even if that would better serve her preferences.²⁹

the upshot of the argument to be that such pairs are (after all) equally *F*, perhaps on the basis that intuition strongly suggests that comparative relations are limited to the standard three. This is often referred to as the Trichotomy Thesis. Do we have reason to think that the Small Unidimensional Difference Principle is *more* intuitive than the Trichotomy Thesis? The point I’m getting at here is well captured by the adage that one person’s *modus ponens* may be another’s *modus tollens*. Thanks to Elia Zardini for helpful comments on this point.

²⁵ Quinn (1990), 90.

²⁶ *Ibid.*, 79.

²⁷ *Ibid.*

²⁸ *Ibid.*, 89.

²⁹ *Ibid.*, 88.

Tenenbaum and Raffman (2012) draw similar conclusions about the shortcomings of orthodox decision theory from the puzzle, but propose a different solution from Quinn's pre-selection strategy (and similar plan-based strategies). The agent, they say, doesn't just care about money and freedom from pain, she cares about making *enough* money and – crucially – about living a *relatively* pain-free life. These attitudes are importantly different from momentary preferences and constitute what Tenenbaum and Raffman call 'vague projects'.³⁰ In particular, in relation to the puzzle, the latter project – on grounds of instrumental necessity – permits the agent to stop turning the dial at some non-specific point within a certain (fuzzy) range. (Since the competing projects are vague, there is no optimal trade-off point between pain and money.) Or else the agent's vague project is not realized. But this requires the agent not to maximize expected utility based on her preferences for available actions 'considered in isolation,' as they put it.³¹ Despite the difference in the proposed solution from Quinn's (there is no pre-selected/planned stopping point), therefore, they identify as the culprit the same aspect of the orthodox view: It requires all the practically relevant aspects of the agent's make-up to figure *directly* in her momentary choices. Vague projects, however, are just not like that, Tenenbaum and Raffman say. The extent to which such projects are accomplished can only be assessed, they claim, 'in light of the entire period during which they were, or ought to have been, executed.'³²

Quinn, and Tenenbaum and Raffman, then, take the self-torturer to show that *rationaly-preferable-to* is a not transitive relation, and think that any complete theory of rationality must adequately address and accommodate this. And while Tenenbaum and Raffman are silent on this, Quinn takes *better-than* to be transitive; plausibly, on Quinn's view, the puzzle is generated in part by this structural difference between these two key relations. For each adjacent pair of settings, there is a significant difference between them in terms of monetary payoff but no perceivable difference in terms of pain, so it is always rational to prefer the higher setting to the lower (current) one. But, since *rationaly-preferable-to* is not transitive there is no inconsistency involved in the self-torturer's preference to, say, prefer setting 0 to setting 1000. The same is not true of *better-than*, Quinn thinks. If a further increase in the setting were always *better*, then setting 1000 *would* – via transitivity – be better

³⁰ Tenenbaum and Raffman (2012), 99–106.

³¹ *Ibid.*, 102.

³² *Ibid.*

than setting 0. Since action should – on Quinn’s view – track goodness rather than merely preference-satisfaction, a complete theory of rational action must (among other things) make up the shortfall when the latter does not align with the former.

Temkin’s continuum argument against the transitivity of better-than

I hope to have made clear how much work the assumption that *better-than* is transitive does in Quinn’s overall argument. Fortunately for Quinn, this is a pretty standard assumption in value theory. Temkin (1996), however, provides what he calls a ‘continuum’ argument (later ‘spectrum’ argument) for the controversial thesis that *better-than*, in the ‘all things considered’ sense, is – or at least *may* be – non-transitive. The argument is ultimately supposed to motivate us to ‘rethink our understanding of the good, moral ideals, and the nature of practical reasoning.’³³

Put very roughly, the dilemma Temkin intends to expose is that – due to arguments like the one below – we will have to either give up some of our most robust ideas about how to evaluate lives or else reject transitivity, which would fundamentally upset what is commonly taken to be our best understanding of practical rationality. The former more or less carries its significance on its sleeve. The latter, among other things, would affect not only many of our common-sensical choice strategies (like choice by elimination) but also many of our most accomplished models in game theory, decision theory, and economics.³⁴ A lot is at stake, then.

Temkin’s continuum argument against the transitivity of *better-than* rests on three fundamental claims, the first and third of which he thinks are deeply entrenched in the way we think about the goodness of outcomes (the second is taken to be a straightforward truism):³⁵

Claim 1: For any unpleasant or ‘negative’ experience, ... it would be better to have that experience than one that was only a little less intense but twice as long.³⁶

³³ Temkin (1996), 210.

³⁴ See e.g. Temkin (2014).

³⁵ Temkin (1996), 179. Temkin’s argument develops a line of argument first introduced in Rachels (1993); see also Rachels (1998). For Temkin’s reason to think that Claims 1-3 are – and should be – claims to which most people are deeply committed, see Temkin (2012), 26–44.

³⁶ This is an instance of what Temkin calls the *First Standard View: Trade-offs between Quality and Number are Sometimes Desirable*; see Temkin (2014), 65. See also Temkin (2012), 30.

Claim 2: There is a continuum of unpleasant or ‘negative’ experiences ranging in intensity, for example from extreme forms of torture to the mild discomfort of a hangnail.

Claim 3: A mild discomfort for the duration of one’s life would be preferable to two years of excruciating torture, no matter the length of one’s life.³⁷

Temkin’s idea is that, starting with excruciating torture for a limited amount of time, decreasing the intensity slightly but the duration significantly will make for a worse state of affairs (Claim 1). However, repeating this pattern of reasoning (Claim 2) will also – assuming that the ‘all things considered better than’ relation is transitive – yield the conclusion that excruciating torture for a limited amount of time is better than very mild discomfort for a vast amount of time (contra Claim 3). The conclusion is that the *all-things-considered-better-than* relation is not transitive, unless – of course – we are willing to give up at least one of the three fundamental claims. Either way, we are required to give up certain fundamental ideas about the nature of goodness and/or practical rationality.

Historically, Temkin has been in favor of rejecting transitivity. One of the more significant upshots of doing so is that it allows us to block certain arguments for what is known as the *Repugnant Conclusion* – roughly, that for any world full of happy people there is another world full of unhappy people which is better simply in virtue of being significantly more populated.³⁸ In recent work, however, Temkin says that rather than counting

³⁷ This is an instance of what Temkin calls the *Second Standard View: Trade-offs between Quality and Number are Sometimes Undesirable Even When Vast Numbers are at Stake*; see Temkin (2014), 65. See also Temkin (2012), 32.

³⁸ See Parfit (1984), 388 and – generally – 381ff. See also Arrhenius and Tännsjö (2017). Rejecting transitivity of course only blocks those arguments that rely on a piecemeal transition from a world, A , with n happy people to a world, Z , with $n+m$ unhappy people, such that – repugnantly – Z is better, or at least no worse, than A (assuming transitivity).

Total utilitarians are stuck with the Repugnant Conclusion irrespectively of any such argument, simply in virtue of the pairwise comparison between A and Z (because the latter contains more total happiness than the former). Average utilitarians, on the other hand, are not affected simply by pairwise comparison, but are – according to Parfit – able to get around the incremental arguments only because the average principle violates what he calls the *principle of mere addition*: roughly, that adding worthwhile lives to A will not make the resulting world, $A+$, worse (and may in fact make it better); see Parfit (1984), 420. But, and here is the rub, if we accept this principle (and make certain other plausible assumptions), as we arguably should if we accept that welfare affects

equivocally in favor of non-transitivity, continuum arguments instead expose us to a difficult dilemma in theorizing about morality and action. '[T]he question of which of the premises should be given up,' he says, 'is a difficult one about which people are deeply divided, and about which there is unlikely to be a consensus for years to come;³⁹ the way forward, he thinks, is 'murky, at best.'⁴⁰

The sorites charge

All three arguments naturally invite the idea that they might be soritical, and thereby fallacious. Temkin, for example, spends a great deal of effort to resist this charge. His main argument is that unlike his continuum argument, sorites reasoning can be rejected once we get clear about how to interpret the relevant tolerance principle. The paradox, on his view, trades on treating insignificant differences as no differences at all. In the case of 'bald', for example, we can either take the relevant principle to be that adding or subtracting one hair will not make a *significant* difference with respect to someone's being bald. But then the tolerance principle – although true – is no longer apt to fuel the paradox. Or we can take the principle to be that adding or subtracting one hair will not make *any* difference whatsoever to someone's being bald, in which case the principle is false. Either way, the sorites reasoning is blocked, Temkin argues. But no such revisionary strategy is available with respect to his continuum argument, he thinks – morality, unlike language, is not a matter of convention. The only way to avoid the conclusion of his argument is to 'revise substantially our understanding of the world,' he says.⁴¹

As for Quinn's self-torturer, Elson (2016) argues that the reasoning involved in the puzzle is an instance of sorites reasoning. Tenenbaum and Raffman think that it isn't, but

goodness in a significant way, then we can – *assuming transitivity* – incrementally rearrange the quality and distribution of happiness such that we eventually conclude that *Z* is better (or at least no worse) than *A*; see Parfit (1984), 419ff.

Temkin's (1996) suggestion is that rejecting transitivity allows us both to block such incremental arguments and to explain why we can accept that, for any (ordered) pair of *adjacent* worlds in the series from *A* to *Z*, the latter is better than then former, while also accepting that *Z* is not better than *A*. Temkin's fundamental idea is that the factors relevant to evaluating the relative goodness of adjacent worlds may be different from the factors relevant to evaluating non-adjacent worlds; see e.g. Temkin (2014), 539. Thanks to the editors for prompting me to clarify this.

³⁹ See Temkin (2014), 64–65.

⁴⁰ *Ibid.*, 58.

⁴¹ Temkin (1996), 201. A number of people have responded to Temkin's claim that his argument is not soritical, some of whom he has replied to in Temkin (2012); see esp. Ch. 9 and Appendices C and D.

provide fairly limited argument to that effect. Primarily, they rely on the fact that the series involved in the self-torturer case is not unidimensional or monadic – the agent isn't faced with a series of indiscriminable steps from bearable to unbearable pain, but with deciding whether each step is compensatable by \$10,000.

Elson argues – contra Tenenbaum and Raffman – that fallacious sorites reasoning is not limited to monadic predicates, and invokes what he calls 'essentially comparative desires,' the function of which is to ground a 'practical sorites' directly in the comparative task itself, rather than in any vague monadic predicate occurring in the description of the relevant desire.⁴²

An essentially comparative desire with respect to some feature F is a desire for things that are more F rather than less F , or vice versa.⁴³ The self-torturer, for example, has a desire for less pain rather than for more pain, and a desire for more money rather than for less money. Given that these generate vague utility functions for the self-torturer, it will be vague at what setting she rationally prefers less pain to more money. This, Elson says, means that we get a sorites on 'maximizes expected utility': Counting upwards from setting 0, there will be a non-definite number of settings that determinately maximize expected utility, counting downwards from setting 1000 a non-definite number of settings that determinately do not maximize expected utility, and somewhere in between those two sets a non-definite number of settings for which it is vague whether they maximize expected utility.⁴⁴ Elson thus agrees with Tenenbaum and Raffman that the self-torturer is not trying to discern at what step the pain has gone from bearable to unbearable, but concludes that she *is indeed* 'proceeding along a sorites series of dial-turns from those with clearly positive marginal utility to those with clearly negative marginal utility, attempting to decide where the positive/required ones end and the negative/impermissible begin.'⁴⁵ It seems, then, that we *can* represent the puzzle of the self-torturer as a sorites on something like 'step n is compensatable by \$10,000,' at least

⁴² Elson (2016), 478–482.

⁴³ *Ibid.*, 482.

⁴⁴ Elson's analysis has the counterintuitive implication that it is not rational for the self-torturer to prefer setting 1000 to setting 999, despite the imperceptible difference in pain between them and the added \$10,000. The explanation, on Elson's analysis, is that – at this point – the disutility of pain will have significantly outweighed the utility of money. As we will see below, I think that – ultimately – this analysis is not available.

⁴⁵ *Ibid.*, 486.

once we get clear about the two essentially comparative desires involved in the application of the predicate. Or so Elson argues.

Elson (2014) also argues that – despite Chang’s argument to the contrary – her chaining argument depends on the sorites-susceptibility of the predicate ‘is (in)comparable with’. Chang’s defense against the sorites charge is that not all arguments that have a sorites-like structure are unsound (she ask us to consider, for example, mathematical induction). Rather, it’s the *vagueness* of the relevant predicate that is doing the paradoxical work. But in the chaining argument, she says, vagueness is doing no such work. If it were, she continues, then it would be true of cases like the one involving Mozart and Michelangelo that one artist is better than the other or they are equally good, although indeterminate which of the three standard relations holds.⁴⁶ But since she finds this analysis implausible, and provides several arguments in support of that view,⁴⁷ she rejects the idea that the chaining argument is soritical.

Elson, however, argues that although Chang provides good reasons to think that the above analysis of Mozart and Michelangelo type cases is implausible, her chaining argument could still be trading on vagueness in the relevant predicate, in particular if – contra Chang – we adopt a view on which the two artists are determinately incomparable (i.e. none of standard comparative predicates applies).⁴⁸ On the suggested view, there are three basic regions in relation to Mozart, separated by fuzzy boundaries: *worseness*, *incomparability*, and *betterness* – inhabited, respectively, by Talentlessi, Michelangelo, and the (hypothetical) SuperMichelangelo. As Elson points out, since Chang’s arguments do not rule out such an analysis, and the analysis involves elements that ground typical instances of the paradox, we do seem to have a plausible view on which the chaining argument is fallacious. The sorites charge sticks, Elson thinks.⁴⁹

⁴⁶ Chang (2002), 665.

⁴⁷ *Ibid.*, 682ff.

⁴⁸ While structurally similar, this is different from the suggestion in fn. 23 – i.e. that Michelangelo and Mozart are *equally good*. My guess is that Elson’s reluctance to embrace that analysis comes from a motivation is to preserve some of Chang’s intuitions about the pair, while still providing a diagnosis on which the argument is soritical.

⁴⁹ See Elson (2014), 564–567.

2. Tolerance and transitivity

In order to make the Sorites charges stick, much hangs on showing that we can construct a plausible tolerance principle for each case. Let me address them in the same order as the arguments were initially presented, starting with Chang's argument for the parity relation.⁵⁰

Chang's argument crucially relies on what she calls the *Small Unidimensional Difference Principle*: *If Talentlessi is comparable with Mozart, then Talentlessi+ is also comparable with Mozart.*⁵¹ Talentlessi+, here, is just what we get by minimally improving Talentlessi along a single dimension of creativity. And so on for Talentlessi++, etc. Apply this principle to the relevant series and we eventually get comparability between Michelangelo and Mozart.

In Quinn's case, and Tenenbaum and Raffman's, the relevant principle can be formulated in the following way: *If step n is compensatable by \$10,000 (or: maximizes expected utility), then step n+1 is compensatable by \$10,000 (or: maximizes expected utility).* Apply this principle to the relevant series and we eventually get to an outcome that leaves the agent worse off than she initially was (i.e. setting 0 is better than setting 1000, even taking into account the financial gain).

Finally, in Temkin's case, the tolerance principle – if there is any – is arguably the following, motivated in large part by Claim 1: *If it would be better to have experience e_1 than e_2 , which is only a little less intense but twice as long than e_1 , then it would be better to have e_2 than e_3 , which is only a little less intense but twice as long as e_2 .* Apply this principle to the relevant series and we eventually conclude that excruciating torture for a limited amount of time is better than very mild discomfort for a vast amount of time.

2.1 Can we reject the tolerance principles?

The most popular strategy for blocking the Sorites Paradox is to reject one of the conditional premises, in which case the respective tolerance principle is false. This is indeed the strategy that Elson opts for in his critique of Tenenbaum and Raffman's argument (which would carry over to Quinn's as well); and although Elson is not explicit about whether the same holds for his critique of Chang's argument, I will assume that he favors the same strategy across different cases. In addition, I will further assume that many would

⁵⁰ In what follows, I take the tolerance principles to be material conditionals.

⁵¹ Chang (2002), 674.

prefer this strategy for Temkin's argument as well, at least insofar as the aim is to show it to be soritical.⁵²

There is a wide variety of accounts on which tolerance principles are false. For our purposes here, however, the details do not matter – the main task is to evaluate whether it is plausible to reject the ones identified above. I will argue that – although not strictly necessary – it *may* make sense to do so in the case of Chang's argument for parity, but not in the case of the other arguments. The reason is that in the former case, the tolerance principle figures as an independent premise in the reasoning, while in latter two cases, the relevant principles in fact *follow from* certain features of the scenarios on which the arguments rely.

Consider the self-torturer's predicament. On both Quinn's and Tenenbaum and Raffman's diagnosis, the main features driving the puzzle are, first, the assumption that the rational agent ought to maximize expected utility at each choice point and, second, the fact that at each point the agent does so by taking the money and up the setting one increment. Thus, it follows from the inherent features of the example that for each antecedent/consequent in (any instance of) the tolerance principle, we can independently verify that it is true. In genuine sorites scenarios, this is not the case, which opens up the possibility of rejecting the relevant principle. But since the tolerance principle here follows from features inherent in the scenario, we cannot reject it without rejecting the underlying features as well. Then the argument succeeds, however.⁵³

A similar analysis applies to Temkin's example. On any theory of the good on which – other things being equal – welfare affects goodness, we have independent grounds for thinking that each antecedent/consequent is true. Consequently, the tolerance principle here also follows from the example itself, unlike the sorites case. If that is correct, then the sorites charge does not stick and we have to reject either that welfare affects goodness in the way exemplified by Claims 1 and/or 3 or that *better-than* is transitive.

Things are different in the case of Chang's Small Unidimensional Difference Principle, since the principle does not follow from the example itself; rather, it is grounded

⁵² For related recent discussions of Temkin's argument, see Nebel (2017) and Pummer (2017). And for suggestions for further reading, see Nebel (2017), fn. 30, and Pummer (2017), fn. 21.

⁵³ Since I take the tolerance principles to be material conditionals, the truth of each antecedent/consequent pair suffices for the truth of the relevant principle. Note that this does not entail that there are no borderline cases for the relevant predicate. It just means that it is not soritical relative to the series on which the argument relies. Thanks to Sergi Oms for prompting me to clarify this.

more or less in the same kinds of concerns that generally motivate tolerance principles involved in straightforward instances of the Sorites Paradox.

It is indeed intuitively plausible that comparability is preserved between each adjacent pair of points in a relevant series, but we have independent reasons to accept the sentences involved *only* in the case of the initial premise – i.e. the first antecedent in the chain. The rest of our rational commitments are then supposed to follow from the application of basic logical operations, namely of successive applications of *modus ponens*. This indeed opens up the possibility of rejecting the relevant principle, as Elson suggests we do, although – as we will see below – this is not necessary for showing that the argument fails.

2.2 Validity and the normativity of consequence

In this final section, I want to try to make it as transparent as possible how the differences above affect the plausibility of analyzing the respective arguments as based on fallacious sorites reasoning. The strategy is to show that even adopting what is arguably the most radical way to block the Sorites Paradox – that of weakening the consequence relation – will not suffice to show that Quinn’s self-torturer puzzle and Temkin’s continuum argument are invalid. The framework – developed in Zardini (2015) and outlined below – invalidates Chang’s argument, however. This, though, does not necessarily mean that it is really a sorites argument, as such – but it does mean that it fails for the same reasons as the paradox. The problem with the argument is, as I have indicated, that – unlike Quinn’s and Temkin’s arguments – we do not have independent reasons to think that each link in the relevant chain is true.

Non-transitive consequence

Zardini (2015) presents a model on which the consequence relation is *non-transitive*.⁵⁴ The model validates a number of important classically valid argument forms – such as *modus ponens* – but invalidates certain forms that are notoriously problematic – like the Sorites Paradox. One of the fundamental ideas in this framework is that consequence is not

⁵⁴ See also Zardini (2008a), (2008b), and (forthcoming). For an introductory overview, see Zardini (this volume).

analyzable in terms of truth preservation,⁵⁵ or indeed the preservation of any designated value – rather, it is a relation between a set of values assigned to the premises and a (possibly different) set of values assigned to the conclusion(s).⁵⁶ Once we are free of the preservation metaphor, Zardini thinks, space opens up for a non-transitive consequence relation (given that the metaphor in effect forces transitivity on us, at least *prima facie*). Further – and more positive – considerations in favor of non-transitivity come from inference scenarios in which the cognizer has insufficient reason to accept the conclusion(s) of an argument she recognizes as classically valid, or perhaps even reasons against doing so, despite having reasons to accept the premises.⁵⁷ Since our logical models ought – on Zardini’s account – to track the normative relation between an argument’s premises and conclusion(s), and since these scenarios are hard to explain away as ‘deviant’, non-transitivist logics are indeed well-motivated, he thinks.

As Zardini points out, we can have either non-inferential or inferential reasons to accept a sentence (or both). In the former case, one is rationally required to treat the relevant sentence as an ‘initial point for further reasoning,’ as he puts it, while in the latter one is rationally required to treat the sentence as a ‘terminal point of acceptance.’⁵⁸ Thus, if one has non-inferential reasons to accept a set of premises, then one has inferential reasons to accept the sentences that follow from it. However, if consequence is non-transitive, one does not by the same token have inferential reasons to accept the consequences of these consequences: inferential reasons to accept a set of sentences do not give rise to inferential reasons to accept other sentences. That is, since the relevant normative relation holds only between non-inferential and inferential reasons, and since coupling two valid argument forms together will not necessarily produce a valid argument form, as it does in transitive logics, one will not have inferential reasons to accept the consequences of what one (merely) has inferential reasons to accept.

We see now how this framework blocks the Sorites Paradox. The paradox is generated by iterated instances of *modus ponens*, each of which is valid, yet chained together they produce an invalid argument. We have non-inferential reasons to accept the first

⁵⁵ Consequence does *guarantee* truth preservation in an informal sense, however; see Zardini (2015), 237–244.

⁵⁶ *Ibid.*, 240.

⁵⁷ *Ibid.*, 251–255.

⁵⁸ *Ibid.*, 256. See also Smith (2014), 196–9 [cited by Zardini].

sentence in the series – canonically, ‘ o_1 is F’ – and thereby have inferential reasons to accept the next – canonically, ‘ o_2 is F’. And so long as we are in the determinate extension of ‘F’, we additionally have (independent, albeit related) non-inferential reasons to accept each successive sentence in the series. But as the series moves on, our non-inferential reasons become weaker and eventually give out. That is, in the series there will – assuming, for ease of exposition, a classical metalanguage – be some last o_i , such that (1) we have non-inferential reasons to accept ‘ o_i is F’, (2) we only have inferential reasons to accept ‘ o_{i+1} is F’, and (3) we have no reason at all to accept ‘ o_{i+2} is F’ – despite the fact we accept that ‘ o_{i+2} is F’ follows from ‘ o_{i+1} is F’.⁵⁹ Thus, we have – on this framework – no reason to accept the paradoxical ‘conclusion’ (since it is in fact *not* a consequence of the set of sorites premises). Paradox avoided.

Which arguments does non-transitivity block?

As we have seen, it is crucial to the above treatment of the Sorites Paradox that our non-inferential reasons to accept a given sentence in the relevant series give out, and eventually terminate, as the series progresses. This also marks an important difference with respect to the arguments we have been considering. As we will see, even on a logical framework on which the consequence relation is weakened in this way (so as to be non-transitive), Quinn’s and Temkin’s arguments remain valid, whereas Chang’s does not. The former two arguments may of course still be problematic in some respects, but I take this to at least show that the arguments are not fallacious in virtue of being soritical.

First, Quinn’s and Temkin’s arguments. As we saw, it is a feature of the relevant examples that, at each point in the relevant series, we have non-inferential reasons to accept that the next step maximizes expected utility or increases welfare, respectively. Thus, unlike in a genuine sorites scenario, *the non-inferential reasons do not give out as the series progresses*. The arguments may thus *look* like chaining arguments, but in fact they are not. Using Zardini’s terminology, we can say that Quinn’s and Temkin’s arguments do not involve ‘soritical

⁵⁹ At o_{i+2} , we have – on this simplified model – reached the borderline region of ‘F’. Note that this model assumes an *agnostic* attitude towards borderline cases. If we abandon this assumption, we may – in borderline cases – have inferential reasons either for accepting ‘ o_{i+2} is F’ or for accepting ‘ o_{i+2} is not-F’ (rather than no reason either way). On such models, we must – on pain of irrationality – accept one or the other (but not both), rather than remain agnostic. See Zardini (this volume), [x](#), for further discussion.

reasons' to accept the progressing series – that is, we do not proceed down the slippery slope *due to* having non-inferential reason to accept a tolerance principle. In fact, it is the other way around – we have inferential reasons to accept the respective tolerance principles because we have non-inferential reasons to accept the antecedent and consequent of each of its instances.

In the case of Chang's argument, things are different. Here, we arguably – contra Elson's suggestion – have non-inferential reasons to accept the tolerance principle (given its intuitive appeal), but lack inferential reasons to accept the (classically valid) conclusion that Michelangelo is comparable with Mozart. On a non-transitive framework, the fact that we have non-inferential reasons to accept that comparability is tolerant with respect to small unidimensional differences gives us inferential reason to accept that Talentlessi+ is comparable with Mozart if we have non-inferential reason to accept that Talentlessi is. But as the series progresses, our non-inferential reasons to accept the relevant consequent/antecedent give out – this much is agreed upon, although it is not part of the scenario in question that our reasons fully terminate, which in turn makes it unclear whether we are dealing with a genuine sorites series. Still, as we get closer in the series to Michelangelo, the weight of our non-inferential reasons to accept the relevant sentences decreases to a point at which we are no longer rationally required to accept the sentence in question. At this point, all that remains are inferential reasons to accept it. And one step further, there is no reason at all to accept the relevant sentence. On this account, then, Chang's chaining argument is invalid and therefore does not succeed in establishing the case for parity.

References

- Alchaurrón, C. and Bulygin, E. 1971. *Normative Systems*. Vienna: Springer.
- Arrhenius, G. and Tännsjö, T. 2017. 'The Repugnant Conclusion,' in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*:
<<https://plato.stanford.edu/archives/spr2017/entries/repugnant-conclusion/>>
- Asgeirsson, H. 2015. 'On the Instrumental Value of Vagueness in the Law,' *Ethics* 125: 425–448.
- Chang, R. 2002. 'The Possibility of Parity.' *Ethics* 112: 659–88.

Constantinescu, C. 2014. 'Moral Vagueness: A Dilemma for Non-Naturalism,' *Oxford Studies in Metaethics*, Vol. 9: 152-185.

Eklund, M. This volume. 'The Incoherentist Solution to the Sorites Paradox,' in S. Oms and E. Zardini (eds.), *The Sorites Paradox*, pp. x–y. Cambridge: Cambridge University Press.

Elson, L. 2016. 'Tenenbaum and Raffman on Vague Projects, the Self-Torturer, and the Sorites,' *Ethics* 126: 474–488.

Elson, L. 2014. 'Heaps and Chains: is the Chaining Argument for Parity a Sorites?' *Ethics* 124: 557–571.

Endicott, T. 2001. *Vagueness in Law*. Oxford: Oxford University Press.

Enoch, D. 2017. 'Non-Naturalistic Realism in Metaethics,' in T. McPherson and D. Plunket (eds.), *The Routledge Handbook of Metaethics*, pp. 29–42. New York: Routledge.

Keefe, R. and Smith, P. 1996. 'Introduction: Theories of Vagueness,' in R. Keefe and P. Smith (eds.), *Vagueness: A Reader*, pp. 1–57. Cambridge: MIT Press.

Kurtsal, I. This volume. 'The Influence of the Sorites Paradox in Metaphysics,' in S. Oms and E. Zardini (eds.), *The Sorites Paradox*, pp. x–y. Cambridge: Cambridge University Press.

Magidor, O. This volume. 'The Epistemicist Solution to the Sorites Paradox,' in S. Oms and E. Zardini (eds.), *The Sorites Paradox*, pp. x–y. Cambridge: Cambridge University Press.

Navarro, P. and Rodríguez, J. 2014. *Deontic Logic and Legal Systems*. Cambridge: Cambridge University Press.

Nebel, J. 2017. 'The Good, the Bad, and the Transitivity of Better Than,' *Nous*, doi:10.1111/nous.12198

Oms, S and Zardini, E. This volume. 'Introduction,' in S. Oms and E. Zardini (eds.), *The Sorites Paradox*, pp. x–y. Cambridge: Cambridge University Press.

Parfit. D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Pummer, T. 2017. 'Spectrum arguments and Hypersensitivity,' *Philosophical Studies*, doi:10.1007/s11098-017-0932-3

Quinn, W. 1990. 'The Puzzle of the Self-Torturer,' *Philosophical Studies* 59: 79–90.

Rachels, S. 1993. *A Theory of Beneficence*. PPE thesis, University of Oxford.

Rachels, S. 1998. 'Counterexamples to the Transitivity of Better Than,' *Australian Journal of Philosophy* 76: 71–83.

Raz, J. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.

- Ridge, M. 2014. 'Moral Non-Naturalism,' in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* <<https://plato.stanford.edu/archives/fall2014/entries/moral-non-naturalism/>>
- Sayre-McCord, G. 2017. 'Moral Realism,' in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*: <<https://plato.stanford.edu/archives/fall2017/entries/moral-realism/>>
- Schoenfield, M. 2015. 'Moral Vagueness is Ontic Vagueness,' *Ethics* 126: 257-282.
- Shafer-Landau, R. 2003. *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Shafer-Landau, R. 1995. 'Vagueness, Borderline Cases and Moral Realism,' *American Philosophical Quarterly* 32: 83-96.
- Shafer-Landau, R. 1994. 'Ethical Disagreement, Ethical Objectivism, and Moral Indeterminacy,' *Philosophy and Phenomenological Research* 54: 331-44.
- Smith, N. 2004. 'Vagueness and blurry sets,' *Journal of Philosophical Logic* 33: 165-235.
- Temkin, L. 1996. 'A Continuum Argument for Intransitivity,' *Philosophy and Public Affairs* 25: 175-210.
- Tenenbaum, S and Raffman, D. 2012. 'Vague Projects and the Puzzle of the Self-Torturer,' *Ethics* 123: 86-112.
- Unger, P. 1979. 'There are no ordinary things,' *Synthese* 4: 117-54.
- Von Wright, G.H. 1963. *Norm and Action*. London: Routledge and Kegan Paul.
- Williams, B. 1995. 'Which Slopes are Slippery,' in *Making Sense of Humanity and Other Philosophical Papers*, pp. 213-223. Cambridge: Cambridge University Press.
- Williamson, T. 1994. *Vagueness*. London: Routledge.
- Zardini, E. Forthcoming. 'First-order tolerant logics,' *The Review of Symbolic Logic*.
- Zardini, E. This volume. 'The Non-Transitivist Solution to the Sorites Paradox,' in S. Oms and E. Zardini (eds.), *The Sorites Paradox*, pp. x-y. Cambridge: Cambridge University Press.
- Zardini E. 2015. 'Breaking the Chains. Following-from and Transitivity,' in C. Caret and O. Hjortland (eds.), *Foundations of Logical Consequence*, pp. 221-275. Oxford University Press, Oxford.
- Zardini, E. 2008a. 'A model of tolerance,' *Studia Logica* 90: 337-368.
- Zardini, E. 2008b. *Living on the Slippery Slope: The Nature, Sources and Logic of Vagueness*. PhD thesis, Department of Logic and Metaphysics, University of St Andrews.